

Optimized Group Sequential Study Designs for Tests of Genetic Linkage and Association in Complex Diseases

Inke R. König,* Helmut Schäfer, Hans-Helge Müller, and Andreas Ziegler

Center for Methodology and Health Research, Institute of Medical Biometry and Epidemiology, Philipps-University of Marburg, Germany

The study of genetic linkage or association in complex traits requires large sample sizes, as the expected effect sizes are small and extremely low significance levels need to be adopted. One possible way to reduce the numbers of phenotypings and genotypings is the use of a sequential study design. Here, average sample sizes are decreased by conducting interim analyses with the possibility to stop the investigation early if the result is significant. We applied optimized group sequential study designs to the analysis of genetic linkage (one-sided mean test) and association (two-sided transmission/disequilibrium test). For designs with two and three stages at overall significance levels of .05 and .0001 and a power of .8, we calculated necessary sample sizes, time points, and critical boundaries for interim and final analyses. Monte Carlo simulation analyses were performed to confirm the validity of the asymptotic approximation. Furthermore, we calculated average sample sizes required under the null and alternative hypotheses in the different study designs. It was shown that the application of a group sequential design led to a maximal increase in sample size of 8% under the null hypothesis, compared with the fixed-sample design. This was contrasted by savings of up to 20% in average sample sizes under the alternative hypothesis, depending on the applied design. These savings affect the amounts of genotyping and phenotyping required for a study and therefore lead to a significant decrease in cost and time.

Introduction

In recent years, analyses of genetic linkage or association between a disease status and genetic markers have increasingly been aimed at the localization of genes underlying complex traits. The defining characteristic of complex traits is that they are not inherited in a classical Mendelian manner attributable to a single genetic locus. Rather, they are typically determined by an interaction of environmental and multiple genetic factors (Thomson 2001). One major difficulty of studying the genetic background of complex traits lies in the assumption that the more genes there are involved, the smaller the contribution of each to the overall genetic effect becomes (Wright 1968). To reliably detect small effect sizes, large samples need to be recruited, phenotyped, and genotyped, leading to more-expensive study designs in terms of time and money.

A second factor that contributes to an increase in necessary sample sizes is the adoption of significance

levels that are adjusted for genomewide multiple testing. To illustrate the impact of this factor, we briefly summarize the ongoing discussion on adequate significance levels for genetic linkage and association studies. Thereafter, we introduce the concept of sequential study designs as a possible way to decrease the required sample sizes for genetic epidemiological studies.

Adjustment of Significance Levels for Genetic Linkage and Association Studies

A frequent strategy in linkage studies is screening of large genomic regions or even the whole genome for areas showing genetic linkage to the disease of interest. In the extreme case, the locus of every gene could be tested for linkage with disease status. If unadjusted significance levels are then applied for every test, this approach leads to a tremendous rate of false-positive results. To guard against this, several strategies have been suggested for the analysis of genetic linkage. Lander and Kruglyak (1995) derived stringent adjusted testwise significance levels for the analysis of affected sib pairs and other study designs. However, their presented critical values have been criticized on the grounds of unrealistic specification of parameters; for example, an infinitely dense marker map is assumed (Curtis 1996; Witte et al. 1996; Morton 1998). An alternative is to apply the traditional criterion for significance of a LOD score of 3.0, first proposed by Morton (1955), approximating a testwise significance level of .0001. On the basis of the for-

Received May 29, 2001; accepted for publication July 9, 2001; electronically published July 26, 2001.

Address for correspondence and reprints: Dr. Hans-Helge Müller or Dr. Andreas Ziegler, Institute of Medical Biometry and Epidemiology, Philipps-University of Marburg, Bunsenstr. 3, 35033 Marburg, Germany. E-mail: muellerh@mail.uni-marburg.de or ziegler@mail.uni-marburg.de

* Formerly: Inke R. Böldeker.

© 2001 by The American Society of Human Genetics. All rights reserved.
0002-9297/2001/6903-0012\$02.00

mulas by Lander and Kruglyak (1995), Ott (1999) showed that this criterion results in a genomewide Bonferroni-corrected significance level of $\sim 5\%$ when a marker spacing of 5 cM is used.

As Risch and Merikangas (1996) discussed, genome-wide association studies based on several markers in every gene—for instance, single-nucleotide polymorphisms (SNPs)—might be more powerful than genome-wide linkage analyses; this approach also calls for an adjustment of the testwise significance levels. As a consequence, Risch and Merikangas (1996) have proposed a correction of the significance level to be used in a single test. They showed that, under the assumption of 100,000 genes in the human genome, with five diallelic markers in each gene to be analyzed, a significance level of 5×10^{-8} gives a probability of 95% for no false-positive results. Restricting the number of tests to be performed to possible candidate genes only still requires correcting the significance level. For example, Crowe (1993) set the number of candidate genes to be analyzed for psychiatric disorders to include all of the $\sim 20,000$ genes expressed in the brain. Under the assumption of five genes correctly associated with the disease, he calculated a significance level of 10^{-5} for a genomewide false-positive rate of 5%.

Any of these corrections are costly. Consider, for example, performing the transmission/disequilibrium test (TDT; Spielman et al. 1993) to test for genetic linkage and association in a sample of affected individuals with their parents. Applying the proposed level of 5×10^{-8} , as opposed to the traditional .05, leads to an approximately fivefold increase in required sample sizes, regardless of the underlying genetic model or effect. Even the more liberal criterion of .0001 requires almost three times the sample size as does the criterion of .05. Hence, to apply study designs with these appropriate significance levels, even greater sample sizes are necessary, thus adding to study cost and time.

Sequential Designs in Genetic Epidemiological Studies

A possible way to reduce the average sample size and thus facilitate investigations of genetic association and linkage of complex diseases is the use of sequential study designs. These have been developed mostly for application in clinical trials, where the benefits both of requiring as few probands (or even patients) as possible and of providing new therapies as early as possible has been most pronounced. In genetic linkage or association studies, formal sequential study designs have been applied only seldom, although Morton (1955) suggested the use of the sequential probability ratio test (Wald 1947) in introducing the traditional LOD statistic. Furthermore, sequential procedures considering the special needs of genetic epidemiological studies have been pro-

posed but lack correct realizations for practical purposes (Böddiker and Ziegler 2001).

Nonetheless, two kinds of sequential proceedings could be applied in genetic epidemiological studies. The first strategy has been put forward, for instance, by Guo and Elston (2000). Their approach involves screening the whole genome by use of widely spaced genetic markers in the first stage. On the basis of statistical test results, those markers that are significant according to a calculated criterion are flanked by a certain number of additional markers in the second stage. Hence, instead of all available markers in the whole sample being genotyped at once, the marker density is sequentially increased in promising genomic areas. This procedure therefore leads to a significant decrease in the number of genotypings necessary.

The second sequential strategy parallels applications in clinical trials. Here, the basic idea is to analyze subsets of probands or patients, in the extreme consisting of single probands at a time. If, after the analysis of one subset, the result is significant or definitely unpromising, the trial is stopped. Otherwise, further subsets of individuals are recruited and included in the study until a clear-cut result is obtained. Thus, in this procedure, the sample size is increased sequentially, and multiple significance tests are performed on increased samples. As the number of probands in the study is decreased on average, the amount of recruiting and phenotyping as well as genotyping is reduced. Following this strategy, Müller and Ziegler (1998) have proposed a group sequential design for the TDT. However, their calculations were based on formulas given by Camp (1997) which have been shown to be incorrect (Camp 1999; Knapp 1999).

Using the same approach as Müller and Ziegler (1998), we focus in the present study on group sequential designs for studies of genetic linkage and association. For the analysis of genetic linkage, we apply the mean test using affected sib pairs (ASPs), which is based on the mean proportion of alleles shared identical by descent (IBD) by ASPs. To study linkage and association simultaneously, we apply the TDT (Spielman et al. 1993) that has been one of the most commonly applied tests over the last years. In its original form, this test examines the transmission of a particular marker allele from heterozygous parents to their affected offspring. Here, requiring large sample sizes is especially problematic, since trios (i.e., parents and an affected offspring) must be recruited and phenotyped.

The aim of the present study is to propose study designs for genetic linkage and association studies that reduce the required average sample sizes. To this end, we apply group sequential study designs to both the mean test and the TDT. For a given number of interim analyses, the proposed designs are optimized with respect to

average sample sizes under the alternative hypothesis and are presented with necessary sample sizes, time points, and critical values for interim and final analyses. The validity of the asymptotic approximation is investigated by Monte Carlo simulation analyses. In addition, the superiority of the sequential designs in comparison with fixed-sample designs with regard to average sample sizes will be presented.

Material and Methods

Mean Test for ASPs

To test for genetic linkage in ASPs, we consider the mean statistic. Knapp (1994) showed that, in the case of a multiplicative mode of inheritance (MOI), the mean statistic leads to the test that is uniformly the most powerful and that, in the case of any MOI, no other test based on sib pairs could be uniformly the most powerful. Multiplicative MOI refers to the inheritance model at a single locus where having two disease alleles instead of one squares the genotypic relative risk (GRR; see also table 1). In the notation of Guo and Elston (2000), the statistic of the mean test for N_{ASP} families of ASPs is defined as

$$T_{\text{mean}} = \frac{\sqrt{2N_{\text{ASP}}} \left(\hat{\pi} - \frac{1}{2} \right)}{\sqrt{\hat{\pi}(1 - \hat{\pi})}},$$

where $\hat{\pi}$ is the proportion of alleles shared IBD by an ASP estimated from the data. The null hypothesis of no linkage is rejected if $T_{\text{mean}} (\text{observed}) \geq z_{1-\alpha}$, with z_{α} being the α fractile of the standard normal distribution. Given π , the expected proportion of alleles shared IBD, linkage can be detected at a power of $1-\beta$, with a fixed total sample size (Guo and Elston 2000) of

$$N_{\text{ASP}} (\text{fixed}) \geq (z_{1-\alpha} + z_{1-\beta})^2 \frac{\pi(1-\pi)}{2(\pi - \frac{1}{2})^2}. \quad (1)$$

TDT for Trios

To test for genetic association and linkage in trios, we use the TDT, which examines the transmission of a particular marker allele from heterozygous parents to their affected offspring (Spielman et al. 1993). In the following, we use the notation and genetic models of Camp (1997). Here, the investigated marker represents the disease locus itself, so that the recombination fraction (θ) is 0. We consider a disease allele A with a frequency p , where $q = 1 - p$. For genotypes comprising 0, 1, or 2 A alleles, the probability of expressing the disease is given by f_0 , f_1 , and f_2 , respectively. GRR_i is given by the increased chance that an individual with i disease alleles

Table 1

Penetrance Functions and Genotypic Relative Risks

	GRR FOR MOI			
	Multiplicative	Additive	Recessive	Dominant
GRR_1	γ	γ	1	γ
GRR_2	γ^2	2γ	γ	γ

expresses the disease, compared with an individual with zero A alleles. For $f_0 \neq 0$, they are defined as $\text{GRR}_1 = f_1/f_0$ and $\text{GRR}_2 = f_2/f_0$. Using a penetrance parameter γ , the penetrances can be written in terms of GRR_1 and GRR_2 for different MOI, as shown in table 1.

On the basis of calculations by Knapp (1999), we define v as the probability that a parent is heterozygous and transmits allele A to the affected offspring and ω as the probability that a parent is heterozygous and transmits the other allele. Hence, the difference $v - \omega$ can be used to express the genetic effect, and we test the null hypothesis that $v - \omega = 0$ against the two-sided alternative hypothesis that $v - \omega \neq 0$.

Camp (1999) has shown that, for the purpose of power calculations, the classical TDT test statistic by Spielman et al. (1993) can be approximated for N_{trio} families with a given difference of $\hat{v} - \hat{\omega}$ by

$$T_{\text{TDT}} = \frac{\sqrt{N_{\text{trio}}}(\hat{v} - \hat{\omega})}{\sqrt{4pq}}.$$

The null hypothesis is rejected if $|T_{\text{TDT}} (\text{observed})| \geq z_{(1-\alpha/2)}$.

The number of families necessary to achieve a power of $1-\beta$ can now be calculated, given the GRRs and allele frequencies, by

$$N_{\text{trio}} (\text{fixed}) \geq \frac{\left[z_{(1-\alpha/2)} \sqrt{2pq} - z_{\beta} \sqrt{S - 2h^2(v - \omega)^2} \right]^2}{2h^2(v - \omega)^2} \quad (2)$$

with h being the probability that a patient of an affected offspring is heterozygous and

$$S = \frac{pq \left[(1-q)^2 \text{GRR}_2 + (1-2pq) \text{GRR}_1 + (1-p^2) \right]}{p^2 \text{GRR}_2 + 2pq \text{GRR}_1 + q^2}.$$

General Group Sequential Procedure

To adopt group sequential designs, we follow the procedure and notation introduced by Müller and Schäfer (1999). We assume that data will be accumulated se-

quentially until the total sample size N_{ASP} of ASPs or N_{trio} of trios is reached. The amount of information at any point is given by the proportion t of families relative to the total sample size and will be termed information time t . Then the test statistics for the mean test and the TDT can be expressed in terms of t with

$$T_{\text{mean}}(t) = \sqrt{t} \frac{\sqrt{2N_{\text{ASP}}} \left(\hat{\pi} - \frac{1}{2} \right)}{\sqrt{\hat{\pi}(1 - \hat{\pi})}}$$

and

$$T_{\text{TDT}}(t) = \sqrt{t} \frac{\sqrt{N_{\text{trio}}}(\hat{v} - \hat{\omega})}{\sqrt{4pq}}$$

and render the cumulative difference in transmission rates up to time t .

For the time parameter t , varying between 0 and 1, $T_{\text{mean}}(t)$ and $T_{\text{TDT}}(t)$ define stochastic processes. They are characterized by mean $= \delta\sqrt{t}$ and unit variance, where

$$\delta_{\text{mean}} = \frac{\sqrt{2N_{\text{ASP}}} \left(\pi - \frac{1}{2} \right)}{\sqrt{\pi(1 - \pi)}} \quad (3)$$

and

$$\delta_{\text{TDT}} = \frac{\sqrt{N_{\text{trio}}} (v - \omega)}{\sqrt{4pq}}. \quad (4)$$

Asymptotically, $\sqrt{t} \cdot T_{\text{mean}}(t)$ and $\sqrt{t} \cdot T_{\text{TDT}}(t)$ follow a Brownian motion with drift parameters δ_{mean} and δ_{TDT} . It can be seen that these drift parameters are determined mainly by the sample size and the genetic effects in terms of π and $v - \omega$. For a one-sided mean test, we now test the null hypothesis H_0 ($\delta_{\text{mean}} \leq 0$) against the alternative H_1 ($\delta_{\text{mean}} > 0$); for a two-sided TDT, we test the null hypothesis H_0 ($\delta_{\text{TDT}} = 0$) against the alternative H_1 ($\delta_{\text{TDT}} \neq 0$). Using formulas (3) and (4), a specific study design can be connected with the Brownian motion; this allows classical sequential designs to be defined according to the procedure of Müller and Schäfer (1999).

α -Spending Approach

To define a general sequential procedure, let m denote the maximum number of analyses in the group sequential design, and $t_1, t_2, \dots, t_m = 1$ are the information times at which analyses are carried out. The k th analysis at information time t_k is performed on the observed value of the test statistic $T(t_k)$. If the number of analyses, m , is fixed, a group sequential design can be determined using the α -spending approach (Lan and DeMets 1983). Here, the subsequent steps are followed:

First, a continuous function $\alpha(t)$ is defined, specifying the type I-error rate that is spent until the information time, t , with $0 \leq \alpha(t) \leq \alpha$. Classical examples for spending functions are $\alpha_p(t) = \alpha \cdot \ln[1 + (e - 1) \cdot t]$, approximating the group sequential designs by Pocock (1977), and

$$\alpha_{\text{OF}}(t) = 4 - 4\Phi \frac{\Phi^{-1}(1 - \frac{\alpha}{4})}{\sqrt{t}},$$

where Φ denotes the standard normal distribution function, corresponding to the designs proposed by O'Brien and Fleming (1979).

The second step includes fixing the testing times t_k for interim and final analyses. Given the respective function and the maximal sample size $N(\text{max})$, the type I error spent at the first analysis is then given by $\alpha(t_1)$ where $t_1 = n_1/N(\text{max})$.

Third, the critical boundary for the first analysis, b_1 , is calculated from $P_{\delta=0}[T(t_1) \geq b_1] = \alpha(t_1)$, rendering $b_1 = \Phi^{-1}[1 - \alpha(t_1)]$ in the case of a one-sided test. The critical boundary for the second analysis, b_2 at $t_2 = (n_1 + n_2)/N_{\text{max}}$, is obtained from $P_{\delta=0}[T(t_1) < b_1 \cap T(t_2) \geq b_2] = \alpha(t_2) - \alpha(t_1)$. To solve for the boundary values, numerical integration methods, as described by Armitage (1969), can be applied, using the independent increments of a Brownian motion. Specification of critical boundaries is continued recursively until the overall significance level is exhausted for $\delta = 0$ under the null hypothesis. The calculation of critical boundaries for a two-sided test is performed analogously.

In the fourth step, the power function can be calculated as a function of the drift parameter δ for a given design specified by t_k and b_k . The value of this function equals α at $\delta = 0$, and a characteristic parameter $\delta = \delta^*$ for the specific study design can be selected, in which the power equals $1 - \beta$. In the next section, it will be shown that δ^{*2} can be interpreted as a factor in the sample-size calculation, characterizing the study design in terms of significance level, spending function, power, and number and time points of interim analyses.

Hence, when the α -spending approach is used, local critical boundaries b_k for each analysis are determined together with δ^* , depending on the spending function $\alpha(t_k)$ and t_k . The result of these calculations is the following statistical test procedure. If the observed statistic $T_{\text{mean}}(k)$ is greater than the critical boundary b_k when the one-sided mean test is performed, the null hypothesis is rejected and the study stopped. Similarly, if the two-sided TDT is performed, the observed statistic $T_{\text{TDT}}(k)$ is compared with the critical values $\pm b_k$, and the null hypothesis is rejected if the observed statistic lies outside the respective boundaries. Otherwise, the study is carried on to the next stage until the last stage, m , in which the null hypothesis is either accepted or rejected.

Table 2
Optimized Group Sequential Designs

Overall α , Test, and m^a	t_1	t_2	t_3	b_1	b_2	b_3
.05:						
One-sided:						
1	1	1.6449
2	.5271	1	...	1.9587	1.7996	...
3	.3990	.6711	1	2.1266	2.0062	1.8502
Two-sided:						
1	1	1.9600
2	.5523	1	...	2.2510	2.1043	...
3	.4257	.6890	1	2.4074	2.2974	2.1528
.0001:						
One-sided:						
1	1	3.7190
2	.6587	1	...	3.9220	3.8226	...
3	.5471	.7641	1	4.0326	3.9632	3.8607
Two-sided:						
1	1	3.8906
2	.6671	1	...	4.0879	3.9908	...
3	.5568	.7697	1	4.1952	4.1281	4.0283

^a $m = 1, 2$, and 3 denote fixed-sample, two-stage, and three-stage group sequential designs, respectively.

Sample-Size Calculation

A prerequisite for using the α -spending approach and for connecting a specific study design with the Brownian motion is the determination of the maximum sample size. Generally, the application of a group sequential design leads to a slight increase in the maximum sample size over the fixed-sample design given in formulas (1) and (2). For a specific study, the required maximum sample size $N(\max)$ can be calculated using the formulas given above for δ_{mean} and δ_{TDT} and the genetic model for the specific study. It can be seen from formulas (3) and (4) that the sample size N is proportional to δ^2 . To calculate the maximum sample size $N(\max)$, equations (3) and (4) are solved for N . For each sequential plan, N is substituted with $N(\max)$, δ with the δ^* for the specific study design, and p and $v-\omega$ with the effects to be detected. This transformation gives

$$N_{\text{ASP}}(\max) = \delta_{\text{mean}}^{*2} \frac{\pi(1-\pi)}{2(\pi-\frac{1}{2})^2}$$

and

$$N_{\text{trio}}(\max) = \delta_{\text{TDT}}^{*2} \frac{4pq}{(v-\omega)^2},$$

which are the required maximum sample sizes to detect the effects π and $v-\omega$ with power $1-\beta$. Note that, in the fixed-sample design, δ^* equals $z_{1-\alpha} + z_{1-\beta}$ in the one-sided case, and $z_{1-\alpha/2} + z_{1-\beta}$ in the two-sided case. The k th interim analysis will be performed with a sample size $N_{\text{ASP}}(k) = t_k \cdot N_{\text{ASP}}(\max)$ and $N_{\text{trio}}(k) = t_k \cdot N_{\text{trio}}(\max)$.

Optimization

Given only type I- and type II-error rates and the maximal number of analyses, a vast number of group sequential plans with different analysis times and critical boundaries can be defined using the above calculations. It is therefore required to select one design that is optimal for the specific study to be conducted. As an optimization criterion, we aim at finding a design that is optimal with regard to average sample sizes. However, since the average sample size depends on the unknown drift parameter, it cannot be specified in advance. To circumvent this uncertainty, we wish to select the design that min-

Table 3
Change in Average Sample Size in the Group Sequential Designs over the Fixed-Sample Design

OVERALL α , TEST, AND m	CHANGE IN AVERAGE SAMPLE SIZE (%)	
	UNDER HYPOTHESIS	
	H_0	H_1
.05:		
One-sided:		
2	+6.47	-16.01
3	+8.37	-20.41
Two-sided:		
2	+5.65	-15.21
3	+7.35	-16.90
.0001:		
One-sided:		
2	+3.68	-11.85
3	+5.02	-15.34
Two-sided:		
2	+3.48	-11.61
3	+4.64	-15.03

Table 4

Number of ASPs N_k for the One-Sided Mean Test for the k th Analysis in the Fixed-Sample Design and in Group Sequential Designs with Two and Three Stages

RELATIVE RISK RATIOS	θ^a											
	0						.05					
	Fixed Sample	Two Stage		Three Stage			Fixed Sample	Two Stage		Three Stage		
	N_1	N_1	N_2	N_1	N_2	N_3	N_1	N_1	N_2	N_1	N_2	N_3
$\lambda_S = 1.2$:												
$\lambda_O = 1.2$	443	252	477	195	327	487	676	384	728	297	499	744
$\lambda_S = 1.5$:												
$\lambda_O = 1.2$	41	23	44	18	30	45	64	36	69	28	47	70
$\lambda_O = 1.5$	109	62	117	48	80	120	167	95	180	74	123	184
$\lambda_S = 2$:												
$\lambda_O = 1.2$	13	7	14	6	9	14	21	12	22	9	15	23
$\lambda_O = 1.5$	19	11	21	9	14	21	31	18	33	14	23	34
$\lambda_O = 2$	47	27	50	21	35	52	73	42	78	32	54	80
$\lambda_S = 3$:												
$\lambda_O = 1.2$	5	3	5	3	4	6	9	5	10	4	7	10
$\lambda_O = 1.5$	6	4	7	3	5	7	11	7	12	5	8	12
$\lambda_O = 2$	10	6	10	5	7	11	16	9	17	7	12	18
$\lambda_O = 3$	25	15	27	11	19	28	40	23	43	18	30	44

^a θ = recombination fraction between diallelic trait and fully informative marker locus. $\alpha = .05$; $\beta = .2$.

minimizes the average sample size under the alternative hypothesis. This optimization problem can be solved by a search algorithm in addition to an algorithm for the α -spending method.

Here, increasing testing times $t_1 < \dots < t_{m-1}$ and cumulative α -spending values $\alpha_1 < \dots < \alpha_{m-1}$ are chosen. The critical boundaries and other characteristics of each design can be calculated, where the average sample size under the alternative hypothesis—the value of the optimization criterion—is of particular importance. Employing a $2(m-1)$ dimensional search, a specific design in terms of testing times and α -spending values can be selected that leads to the minimal value of the optimization criterion.

Using this approach, optimized study designs were determined for the analysis of ASPs and trios with affected offspring. For illustration, designs with maximally two and three stages were calculated for an overall significance level of .05 and a power of .8. Additionally, designs with an overall significance level of .0001 were determined. Since the mean test is performed as a one-sided test and the TDT as a two-sided test corresponding to the original form, both one-sided and two-sided designs are presented for different genetic models.

Monte Carlo Simulations

Monte Carlo simulations were used to evaluate the resulting conventional fixed-sample and sequential designs. Within each replication, a sample of the calculated sample size was created. For linkage designs, the cumulative probabilities for sharing 0, 1, or 2 alleles IBD are given by Risch (1990). By use of these cumulative

probabilities and uniformly distributed random numbers, each ASP was assigned to share 0, 1, or 2 alleles IBD. The proportion of alleles shared IBD in the whole sample was then determined to calculate the test statistic. Similarly for the TDT, cumulative probabilities for all possible family types in trios are presented by Knapp (1999). These were used together with uniformly distributed random numbers to assign each trio to a family type. From the frequencies of the different family types, the proportion of heterozygous parents transmitting the disease or any other allele was determined for the calculation of the test statistic.

The number of replications was chosen to estimate all parameters with a confidence of .95 at appropriate precision. Accordingly, the simulation of the power of .8 and of the significance level of .05 was estimated at a precision of $\pm .001$; for the simulation of the significance level of .0001, the precision was set to $\pm .00001$. As a result, for designs with an overall $\alpha = .05$, we simulated 200,000 replications for each model under the null hypothesis and 700,000 replications under the alternative hypothesis. For designs with $\alpha = .0001$, we simulated 3,900,000 replications for each model under the null hypothesis and, again, 700,000 replications under the alternative hypothesis.

Results

Resulting Study Designs

The conventional fixed-sample design and the optimized group sequential designs are presented in table 2 for the different maximal number of stages m . For both

Table 5

Number of ASPs N_k for the One-Sided Mean Test for the k th Analysis in the Fixed-Sample Design and in Group Sequential Designs with Two and Three Stages

RELATIVE RISK RATIOS	θ^a											
	0						.05					
	Fixed Sample	Two Stage		Three Stage			Fixed Sample	Two Stage		Three Stage		
	N_1	N_1	N_2	N_1	N_2	N_3	N_1	N_1	N_2	N_1	N_2	N_3
$\lambda_S = 1.2$:												
$\lambda_O = 1.2$	1,488	1,016	1,542	854	1,193	1,560	2,273	1,552	2,356	1,304	1,822	2,384
$\lambda_S = 1.5$:												
$\lambda_O = 1.2$	136	93	141	78	109	143	213	146	221	122	171	223
$\lambda_O = 1.5$	364	249	378	209	292	382	561	383	581	322	450	588
$\lambda_S = 2$:												
$\lambda_O = 1.2$	41	28	43	24	33	43	68	47	71	39	55	72
$\lambda_O = 1.5$	64	44	66	37	51	67	103	70	107	59	83	108
$\lambda_O = 2$	156	107	162	90	126	164	244	167	253	140	195	256
$\lambda_S = 3$:												
$\lambda_O = 1.2$	16	11	17	9	13	17	30	20	31	17	24	31
$\lambda_O = 1.5$	21	14	21	12	17	22	37	25	38	21	30	38
$\lambda_O = 2$	32	22	33	18	26	33	54	37	55	31	43	56
$\lambda_O = 3$	84	57	87	48	67	88	133	91	138	76	107	139

^a θ = recombination fraction between diallelic trait and fully informative marker locus. $\alpha = .0001$; $\beta = .2$.

significance levels α , analysis times t_1 , t_2 , and t_3 and critical boundaries b_1 , b_2 , and b_3 are given for all analyses.

To illustrate the application, a two-sided test design with three stages and an overall significance level of .0001 is considered. The first interim analysis is performed on 55.68% of the maximum sample size. The test statistic of the first interim analysis, $T(.5568)$, is compared with the critical boundary b_1 . If $T(.5568)$ is >4.1952 or <-4.1952 , the null hypothesis is rejected for significance. However, if $T(.5568)$ falls within the critical boundaries, the next 21.29% of the families are recruited and genotyped. For the second interim analysis, the cumulative test statistic on 76.97% of the maximum sample size, $T(.7697)$, is then compared with the critical boundary b_2 . If $T(.7697)$ lies outside the critical boundaries of ± 4.1281 , the null hypothesis is rejected. Otherwise, the remaining 23.03% of the sample are included in the study, and the final analysis is performed using the critical boundary b_3 . If the cumulative test statistic $T(1)$ lies outside the critical boundaries of ± 4.0283 , the null hypothesis is finally rejected; otherwise, it is accepted.

Calculated Sample Sizes

To give an overview of the cost and savings of the group sequential designs in comparison with fixed-sample designs, table 3 presents the calculated change in required sample size. For each sequential design, the percent increase or decrease in sample size over the fixed-sample design is given.

The above example of a two-sided study design with

a maximum of three stages at $\alpha = .0001$ is considered for illustration. If the null hypothesis is true, the group sequential design requires, on average, 104.64% of the sample size of the fixed-sample design. In contrast to that, under the alternative hypothesis, the average sample size is reduced by 15.03% in the group sequential designs, compared with that in the fixed-sample design. In a given study, the sample size can even be reduced by 44.32%, as is shown in table 2.

Overall, it can be seen from table 3 that the cost, in terms of sample size, under the null hypothesis is always outweighed by at least twice the savings in mean sample size under the alternative hypothesis. Even greater reductions are expected in cases where the underlying genetic effect exceeds the one assumed for the alternative hypothesis. Both additional cost and savings rise with the maximal number of stages in the design and with the overall significance level in the given designs.

Sample sizes were computed for the one-sided mean test for designs with a fixed sample and for group sequential designs with two and three stages. Varying genetic effects in terms of relative risk ratios for parent-offspring pairs and for sib pairs— λ_O and λ_S , respectively—and for recombination fractions θ were assumed. Tables 4 and 5 present the necessary sample sizes for all interim and final analyses for the mean test for designs with overall significance levels of .05 and .0001, respectively. Further sample sizes not listed in the tables are available upon request.

Analogously, we calculated sample sizes for the conventional fixed-sample design and for the optimized group sequential designs with a maximum of two and

Table 6

Number of Trios N_k for the Two-Sided TDT for the k th Analysis in the Fixed-Sample Design and in Group Sequential Designs with Two and Three Stages

γ AND p	MOI ^a											
	Multiplicative						Additive					
	Fixed Sample	Two Stage		Three Stage			Fixed Sample	Two Stage		Three Stage		
	N_1	N_1	N_2	N_1	N_2	N_3	N_1	N_1	N_2	N_1	N_2	N_3
$\gamma = 1.5$:												
$p = .05$	1,982	1,369	2,051	1,155	1,597	2,074	1,727	1,193	1,787	1,007	1,391	1,807
$p = .1$	1,098	758	1,136	640	885	1,149	852	589	882	497	687	892
$p = .5$	560	387	580	327	452	586	275	190	284	161	222	288
$p = .8$	1,098	758	1,136	640	885	1,149	494	341	511	288	398	516
$\gamma = 2$:												
$p = .05$	520	359	538	303	419	544	520	359	538	303	419	544
$p = .1$	302	209	312	176	244	316	302	209	312	176	244	316
$p = .5$	202	140	209	118	163	211	202	140	209	118	163	211
$p = .8$	454	314	470	265	366	475	454	314	470	265	366	475
$\gamma = 3$:												
$p = .05$	143	99	148	84	116	150	163	113	168	95	131	170
$p = .1$	90	63	93	53	73	94	113	79	117	66	91	118
$p = .5$	90	63	93	53	73	94	152	105	157	89	123	159
$p = .8$	237	164	245	139	191	248	419	289	433	244	338	438
$\gamma = 4$:												
$p = .05$	70	49	72	41	57	73	86	60	89	51	70	90
$p = .1$	47	33	49	28	38	49	68	47	70	40	55	71
$p = .5$	63	44	65	37	51	66	133	92	137	78	107	139
$p = .8$	180	125	187	106	146	189	402	278	416	235	325	421

^a $\alpha = .0001$; $\beta = .2$; $\theta = 0$.

three stages for the two-sided TDT with overall significance levels of $\alpha = .05$ and $.0001$ at a power of 80%. The results for designs with $\alpha = .0001$ are given in tables 6 and 7 for varying genetic effects, allele frequencies, and MOI. Again, further numbers are available from the authors.

Simulated Type I- and Type II-Error Rates

Using the optimized designs with respective sample sizes, Monte Carlo simulations were performed to verify type I- and type II-error rates. Various sets of model parameters, in terms of genetic effect and overall significance level, with sample sizes of up to 1,500 ASPs or trios in the fixed-sample design, were simulated. Except for small sample sizes (<100 families), the asymptotic approximation leads to valid results, and simulated and proposed error levels match well, regardless of the maximal number of stages in the design (results not shown). Furthermore, it should be noted that the asymptotic approximation yields results for the newly proposed group sequential designs similar to those for the conventional fixed-sample study design.

Simulated Sample Sizes

Within each replicate of the Monte Carlo simulation, we registered the sample size required to reach a con-

clusion in the group sequential designs. Thus, average sample sizes needed in the group sequential designs were calculated and compared with the sample sizes of the fixed-sample designs. Across all simulated designs with varying genetic effects, the simulated sample sizes match remarkably well with those expected from table 3.

For illustration, we consider the one-sided mean test at an overall significance level of $\alpha = .0001$. Figure 1 displays the simulated sample sizes in the fixed-sample design (*gray line*), in the two-stage group sequential design (*solid black line*), and in the three-stage (*dotted black line*) group sequential design, plotted against the sample size in the fixed-sample design.

Since the study designs were optimized with regard to minimizing the average sample size under the alternative hypothesis, sample sizes under the null hypothesis are always greater than in the fixed-sample design. If, for example, the fixed sample requires 800 ASPs, the two-stage design needs ~830 ASPs under the null hypothesis, and the three-stage design ~840 ASPs. In contrast to that, the average sample sizes under the alternative hypothesis are much smaller than in the fixed-sample design. In our example, the two-stage design requires only ~700 ASPs, on average, and the three-stage design only ~680 ASPs, compared to the 800 ASPs necessary for the fixed-sample design under the alternative hypothesis. Hence, there is also a visible difference between the two-

Table 7

Number of Trios N_k for the Two-Sided TDT for the k th Analysis in the Fixed-Sample Design and in Group Sequential Designs with Two and Three Stages

γ AND p	MOI ^a											
	Recessive						Dominant					
	Fixed Sample	Two Stage		Three Stage			Fixed Sample	Two Stage		Three Stage		
	N_1	N_1	N_2	N_1	N_2	N_3	N_1	N_1	N_2	N_1	N_2	N_3
$\gamma = 1.5$:												
$p = .05$	756,206	521,997	782,498	440,569	609,077	791,307	2,299	1,588	2,379	1,340	1,852	2,405
$p = .1$	100,526	69,393	104,022	58,568	80,969	105,193	1,474	1,018	1,525	859	1,187	1,542
$p = .5$	1,814	1,253	1,877	1,058	1,462	1,899	2,710	1,871	2,804	1,579	2,183	2,836
$p = .8$	1,525	1,053	1,578	889	1,228	1,595	30,658	21,163	31,724	17,862	24,694	32,081
$\gamma = 2$:												
$p = .05$	189,524	130,826	196,114	110,418	152,650	198,321	630	435	652	367	508	659
$p = .1$	25,383	17,522	26,265	14,789	20,445	26,561	436	301	451	254	351	456
$p = .5$	560	387	580	327	452	586	1,098	758	1,136	640	885	1,149
$p = .8$	589	407	609	343	475	616	13,442	9,280	13,910	7,832	10,827	14,066
$\gamma = 3$:												
$p = .05$	47,618	32,871	49,274	27,743	38,354	49,828	187	129	193	110	151	196
$p = .1$	6,472	4,468	6,697	3,771	5,214	6,773	147	102	152	86	119	154
$p = .5$	202	140	209	118	163	211	560	387	580	327	452	586
$p = .8$	285	197	295	166	230	298	7,459	5,149	7,718	4,346	6,008	7,805
$\gamma = 4$:												
$p = .05$	21,269	14,682	22,009	12,392	17,132	22,257	97	68	101	57	79	102
$p = .1$	2,934	2,026	3,036	1,710	2,364	3,070	85	59	88	50	69	89
$p = .5$	122	85	127	72	99	128	421	291	436	246	340	441
$p = .8$	208	144	215	121	168	217	5,853	4,041	6,057	3,411	4,715	6,125

^a $\alpha = .0001$; $\beta = .2$; $\theta = 0$.

and the three-stage designs, with the three-stage design leading to even lower average sample sizes under the alternative hypothesis. Clearly, these savings outweigh the cost in sample size under the null hypothesis.

Discussion

As is outlined in the Introduction, the requirement of large sample sizes presents a major obstacle for genetic studies of complex diseases. This is mainly due to possibly small genetic effects to be detected and to the necessary adjustment of significance levels in genomewide studies. In view of this problem, it has been emphasized by Gu and Rao (2001) and Terwilliger and Göring (2000) that the optimization of study designs plays a critical role in successful mapping of complex diseases.

Accordingly, the aim of this article is to propose the application of group sequential study designs that minimize the average sample size under the alternative hypothesis. We present optimized designs for linkage and association studies, using the mean test for ASPs and the TDT for trios. Our simulation results have affirmed that these designs lead to asymptotically valid type I- and type II-error levels.

Most importantly, our calculations have revealed significant reductions in average sample sizes under the

alternative hypothesis. The demonstrated savings affect the amount of both phenotyping and genotyping of families necessary for the study. Hence, they lead to a tremendous decrease in expected study cost. This enhances the feasibility of large-scale studies in which multiple hypotheses are tested and the significance levels are adjusted accordingly. Although our design does not eliminate the need for these adjustments, the impact of the resulting increase in sample size is weakened. Furthermore, when sequential designs are used, results are obtained faster than with a fixed-sample design. If statistical analyses are performed sequentially, but the actual sampling is not, it is possible that families can be saved for later independent studies (Province 2000). Clearly, these advantages make group sequential designs an extremely interesting tool for optimization of study designs for the genetic background of complex diseases.

The presented reductions by group sequential designs are expected to be even greater if strictly sequential study designs like the sequential probability ratio test (Wald 1947) are used instead of group sequential designs. However, these are difficult to perform in practice, since they require constant monitoring and statistical analyses after the genotyping of each family.

The reader should note that the application of the presented study designs is restricted by our specific model assumptions. To be precise, the designs for the

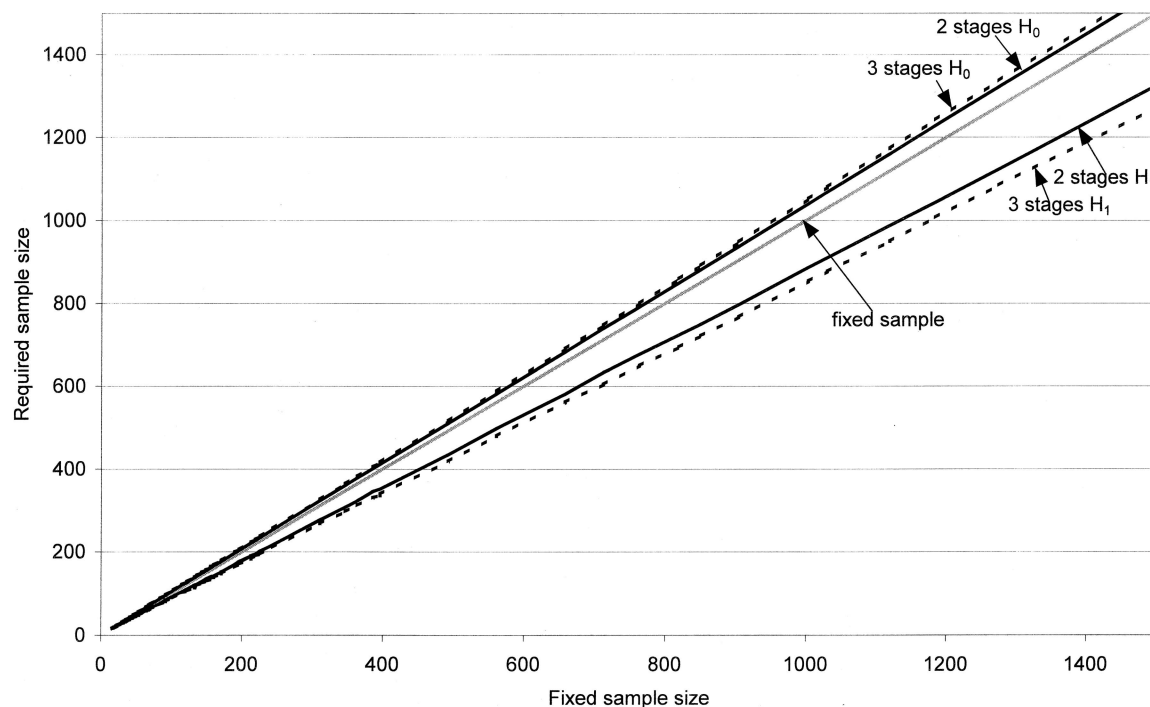


Figure 1 Simulated average sample sizes for the one-sided mean test under H_0 and H_1 . $\alpha = .0001$, and $\beta = .2$; number of replications = 3,900,000 under H_0 and 700,000 under H_1 . *Gray line*, simulated average sample sizes in the fixed-sample design. *Solid and dotted black lines*, simulated average sample sizes in the group sequential designs with two and three stages, respectively.

one-sided mean test require fully informative marker loci. For less-than-fully-informative markers, adjustments of the sample-size calculation have been derived (e.g., Guo and Elston 2000). In the development of sequential designs for the two-sided TDT, we regarded only the genetic models considered by Risch and Merikangas (1996) and by Camp (1997). Thus, it is assumed that the marker locus is identical to the disease locus, and developments are necessary to include different models.

Several modifications or extensions to the presented procedures might further improve the benefit of our designs. First, we only consider an early stopping for significance—that is, the study can be terminated after an interim analysis only if the null hypothesis is rejected. An interesting extension of this would be the inclusion of the possibility of early stopping for futility. In that case, the study can also be terminated after an interim analysis if the null hypothesis is accepted because of a low probability for a significant result, even if the study was carried on. This extension might be especially helpful in genomewide analyses where an early stopping for futility could occur for large genomic areas. Later stages could then be performed only in genomic areas with some probability of significant evidence. Our present research efforts focus on the realization of these possibilities.

Second, the tabulated designs might be modified to better meet the practical demands of genetic applications. In most genetics labs, genotyping is performed using 48-, 96-, or even 384-well plates, and it would be more convenient if sample sizes to be genotyped at a time were adjusted to these block sizes.

A final possible improvement is a combination of the presented group sequential design with the sequential designs proposed by Guo and Elston (2000; modified by Ziegler et al. 2001). Instead of increasing the sample size in later stages, they sequentially increase the marker density in interesting chromosomal regions. A combined strategy might be to increase both sample size and marker density only for promising regions.

Possibly including future extensions, group sequential designs have the power to greatly facilitate large-scale genetic epidemiological studies. They have the advantage of reducing study cost and time while simultaneously being easy to implement. Thus, they present an eminently useful approach to optimization of study designs on the genetics of complex diseases.

Acknowledgments

We are grateful to Ralf Kreß for his programming help. This work was supported by the Deutsche Forschungsgemeinschaft.

References

- Armitage P, McPherson CK, Rowe BC (1969) Repeated significance tests on accumulating data. *J Roy Stat Soc Ser A* 132:235–244
- Böddiker IR, Ziegler A (2001) Sequential designs for genetic epidemiological linkage or association studies: a review of the literature. *Biometrical J* 43:501–525
- Camp NJ (1997) Genomewide transmission/disequilibrium testing—consideration of the genotypic relative risks at disease loci. *Am J Hum Genet* 61:1424–1430
- (1999) Genomewide transmission/disequilibrium testing: a correction. *Am J Hum Genet* 64:1485–1487
- Crowe RR (1993) Candidate genes in psychiatry: an epidemiological perspective. *Am J Med Genet* 48:74–77
- Curtis D (1996) Genetic dissection of complex traits. *Nat Genet* 12:356–358
- Gu C, Rao DC (2001) Optimum study designs. *Adv Genet* 42: 439–457
- Guo X, Elston RC (2000) Two-stage global search designs for linkage analysis I: use of the mean statistic for affected sib pairs. *Genet Epidemiol* 18:97–110
- Knapp M (1999) A note on power approximations for the transmission/disequilibrium test. *Am J Hum Genet* 64: 1177–1185
- Knapp M, Seuchter SA, Baur MP (1994) Linkage analysis in nuclear families. 1: optimality criteria for affected sib-pair tests. *Hum Hered* 44:37–43
- Lan KK, DeMets DL (1983) Discrete sequential boundaries for clinical trials. *Biometrika* 70:659–663
- Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 11:241–247
- Morton NE (1955) Sequential tests for the detection of linkage. *Am J Hum Genet* 7:277–318
- (1998) Significance levels in complex inheritance. *Am J Hum Genet* 62:690–697
- Müller H-H, Schäfer H (1999) Optimization of testing times and critical values in sequential equivalence testing. *Stat Med* 18:1769–1788
- Müller H-H, Ziegler A (1998) Sequential testing using the transmission-disequilibrium test. In: Greiser E, Wischnewsky M (eds) *Medizinische Informatik, Biometrie und Epidemiologie—GMDS '98*, vol 83. MMV Medien & Medizin Verlag, München, pp 467–470
- O'Brien PC, Fleming TR (1979) A multiple testing procedure for clinical trials. *Biometrics* 35:549–556
- Ott J (1999) *Analysis of human genetic linkage*. Johns Hopkins University Press, Baltimore
- Pocock SJ (1977) Group sequential methods in the design and analysis of clinical trials. *Biometrika* 64:191–199
- Province MA (2000) A single, sequential, genome-wide test to identify simultaneously all promising areas in a linkage scan. *Genet Epidemiol* 19:301–322
- Risch N (1990) Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *Am J Hum Genet* 46:229–241
- Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. *Science* 273:1516–1517
- Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 52:506–516
- Terwilliger JD, Göring HHH (2000) Gene mapping in the 20th and 21st centuries: statistical methods, data analysis, and experimental design. *Hum Biol* 72:63–132
- Thomson G (2001) Significance levels in genome scans. *Adv Genet* 42:475–486
- Wald A (1947) *Sequential analysis*. John Wiley, New York
- Witte JS, Elston RC, Schork NJ (1996) Genetic dissection of complex traits. *Nat Genet* 12:355–356
- Wright S (1968) *Evolution and the genetics of populations*. Vol 1: Genetic and biometric foundations. University of Chicago Press, Chicago
- Ziegler A, Böddiker I, Geller F, Guo X, Müller H-H (2001) On the total expected study cost in two-stage genome-wide search designs for linkage analysis using the mean test for affected sib pairs. *Genet Epidemiol* 20:397–400